Statistics, Probability, Distributions, & Error Propagation

James R. Graham

2014/9/2

Probability

Probability: Frequentist vs. Evidential

- Frequentist/Objective/Physical probabilities occur in random physical systems such flipping coins or unstable nuclei
  - In experiments yielding a given outcome–a coin landing "heads"–tends to occur at a persistent rate or frequency in long run trials
  - Frequentist or Physical probability makes sense only when dealing with well defined random experiments

Probability: Frequentist vs. Evidential

Evidential/Bayesian/Subjectivist probability

- Can be assigned to any situation with incomplete knowledge, including non-random ones
  - “What’s the chance this kangaroo is left handed?”
- Evidential probability represents the degree to which a statement is supported by evidence.
- Evidential probabilities are “degrees of belief,” defined in terms of inclination to place bets at certain bookmakers odds.

Probability-Classical Definition

"The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible."

Pierre-Simon Laplace (1749 –1827), *A Philosophical Essay on Probabilities*

Probability-Classical Definition

If a random trial can result in N mutually exclusive and equally likely outcomes, and if N_A of these outcomes result in the occurrence of event A, then according to Laplace $p(A) = N_A/N$

- There are problems with Laplace's definition.
  1. Enumeration: applicable only when there is only a finite number of possible outcomes. Some random experiments, e.g., tossing a coin until it comes up heads, yields an infinite set of outcomes.
  2. Circularity: must determine in advance that all possible outcomes are equally likely, e.g., by symmetry arguments, without relying on the concept of probability

Probability: Frequentist

Frequentists propose that the probability of an event is its relative frequency over time, *i.e.*, it is the relative frequency of occurrence after repeating an experiment many of times under identical conditions

- If the number of occurrences, N_A, of an event A in N trials, then

$$p(A) = \lim_{N \to \infty} \left(N_A / N \right)$$

Probability: Frequentist

Infinite repetitions of an experiment are impossible. With finite experiments, different frequencies will appear and the measured probability is different for each trial.

- Probability can only be measured with some error. The measurement error can has to be expressed in probabilistic terms!
- The frequentist's definition is circular

Cox's Axioms

Richard Cox (1946) formulated a self-consistent set of postulates that justify the "logical" interpretation of probability

1. Probabilities are transitive—if p(A) > p(B) and p(B) > p(C) then p(A) > p(C)—ranking is achieved naturally if probability is a real number associated with a proposition
2. If we specify how much we believe a statement is true we implicitly state how much we believe it is false.
3. If we state how much we believe X is true, then state how likely Y is true given X, then we must implicitly have specified how much we believe X and Y are true (conditional probability).

Mathematically:

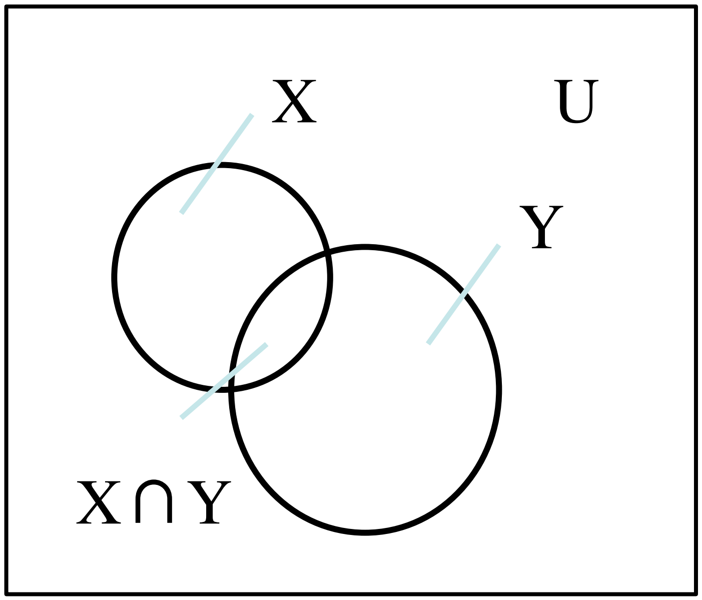
$$p(X) + p(\bar{X}) = 1$$

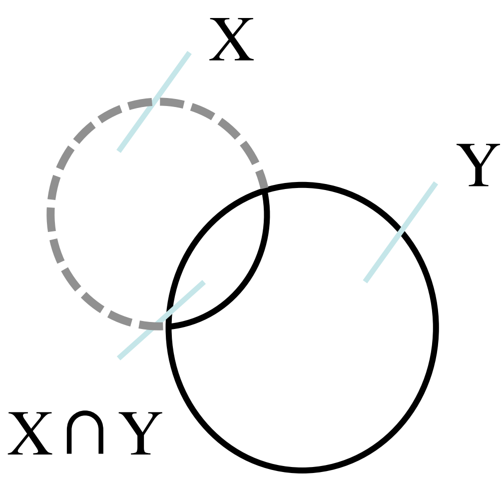
$$p(X \cap Y) = p(Y \mid X) \times p(X)$$

Conditional Probabilities

Suppose we know that event B occurs, then we only need to consider the universe of events which include B and we can exclude the part of A the lies outside of B hence

$$P(X|Y) = P(A \cap Y) / P(Y)$$





The conditional probability P(B|A) is proportional to the joint probability P(A∩B) but it has been rescaled by a factor of 1/P(B) so that the probability in this reduced universe is normalized to 1.

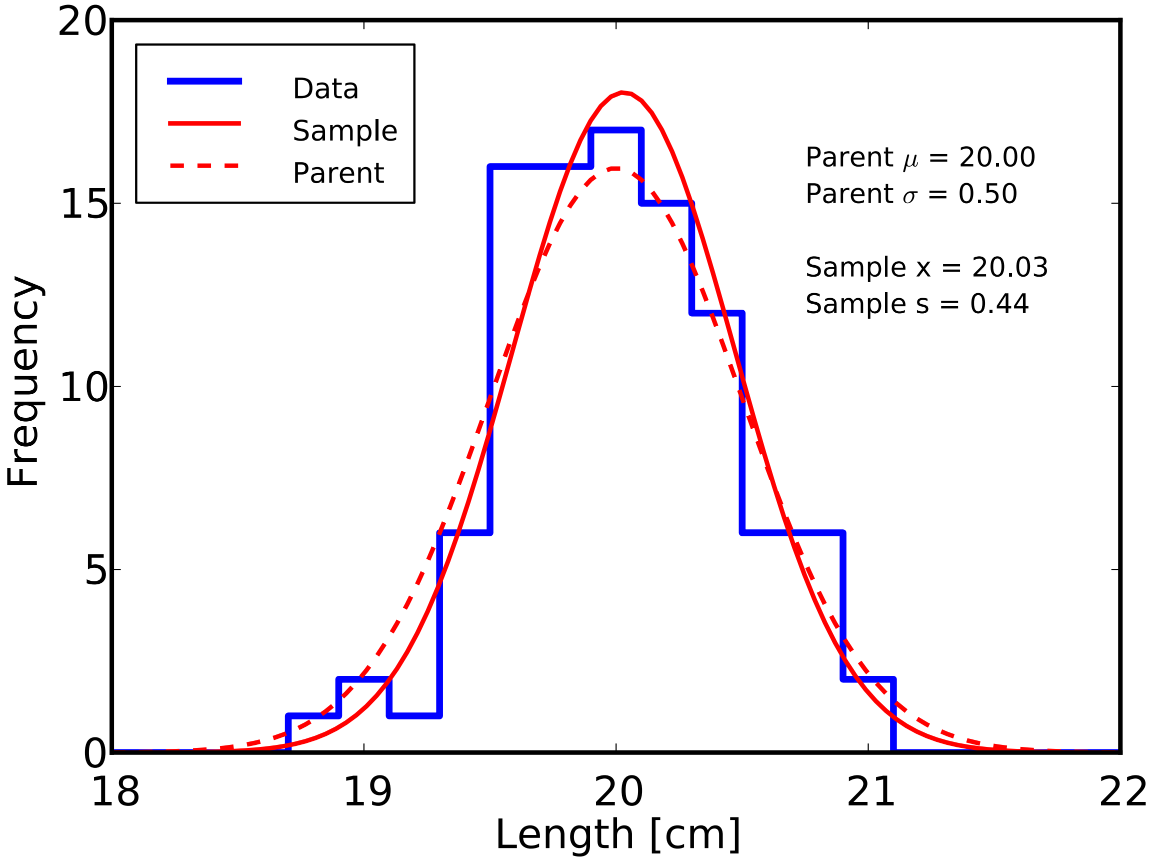
Descriptive Statistics & Probability Density Functions

Sample & Parent Populations

- Make measurements
  - x_1, x_2, …
  - In general $x_1 \neq x_2$
  - With more & more measurements trends emerges
- With an *infinite* sample x_i, i $\in$ {1,2,3…∞} we can:
  - Expect a pattern to emerge with a characteristic value (mean) and dispersion or spread (standard deviation)
  - Exactly specify the ***distribution*** of x_i
- The hypothetical pool of all possible measurements is the ***parent population***
  - Any finite sequence is the ***sample population***

Histograms & Distributions

- A histogram represents the ***occurrence*** or ***frequency*** of discrete measurements
- E.g., length
  - Inferred parent distribution (solid)
  - Parent population (dotted)



Probability Density Functions

- Definition of $P(x)$
  - Limit as $N \to \infty$
  - The number of observations dN that yield values between x and $x + dx$ is

$$dN/N = P(x)\,dx$$

P is usually called the *probability density function* to reflect that it is a differential quantity

Some Useful PDFs

- Binomial

$$P(x;n,p)=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

- Poisson

$$P(x;\mu)=\frac{\mu^x}{x!}\exp(-\mu);\quad x\in\{0,1,2,..\}$$

- Exponential

$$P(t;\tau)=\frac{1}{\tau}\exp(-t/\tau);\quad t>0$$

- Normal/Gaussian

$$P(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-(x-\mu)^2/2\sigma^2\right]$$

Some Important PDF Properties

- Positivity

$$P(x) \geq 0$$

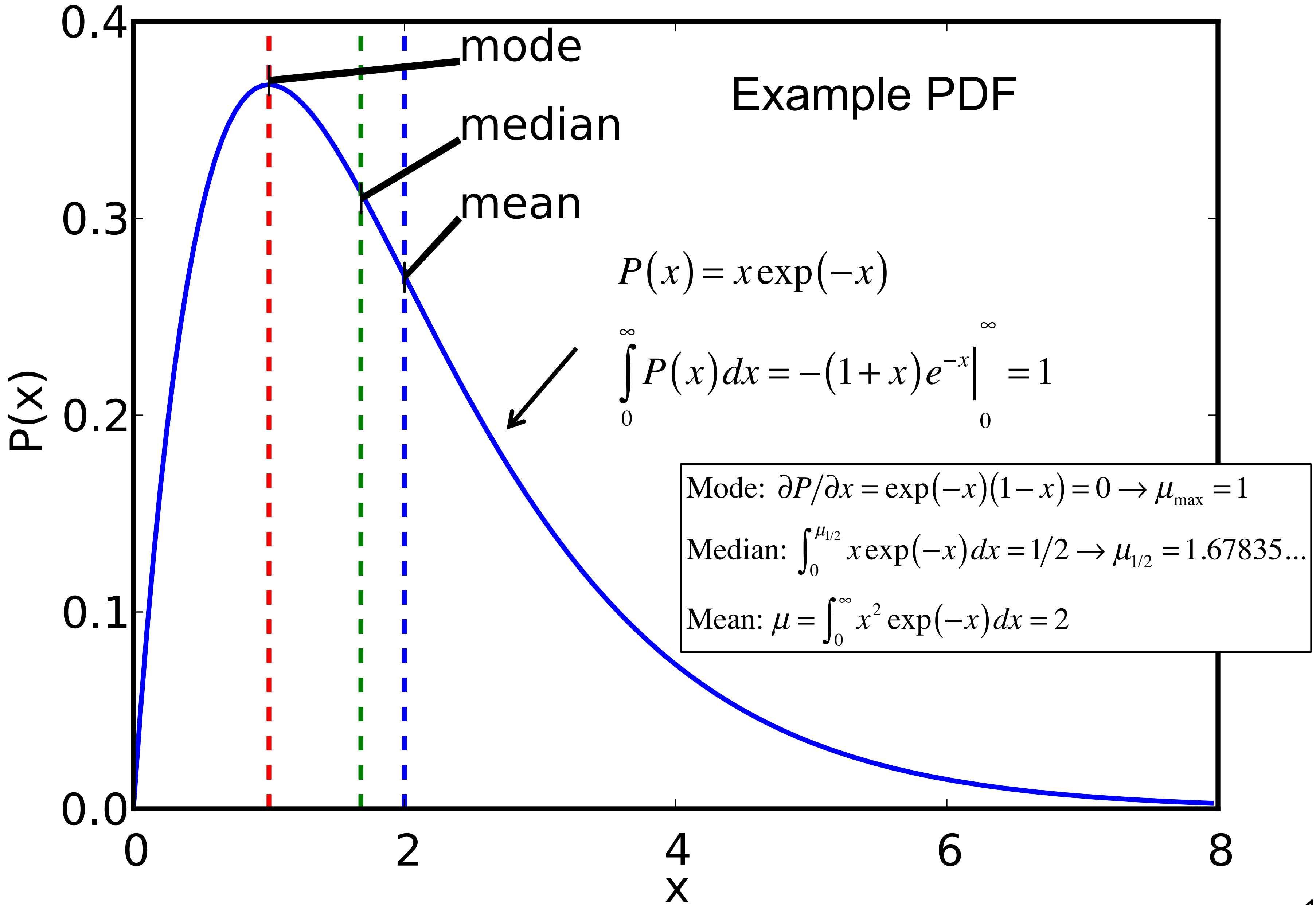
- Normalization

$$\int_x dP(x) = \int_x P(x)\,dx = 1$$

- Moments/expectation values

$$\langle x \rangle = \int_x xP(x)\,dx \qquad \langle x^n \rangle = \int_x x^n P(x)\,dx$$

e.g., variance $\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$


Example PDF
mode
median
mean
$P(x) = x\exp(-x)$
$\int_0^\infty P(x)\,dx = -(1+x)e^{-x}\Big|_0^\infty = 1$
Mode: $\partial P/\partial x = \exp(-x)(1-x) = 0 \rightarrow \mu_{\max} = 1$
Median: $\int_0^{\mu_{1/2}} x\exp(-x)\,dx = 1/2 \rightarrow \mu_{1/2} = 1.67835\ldots$
Mean: $\mu = \int_0^\infty x^2\exp(-x)\,dx = 2$
P(x)
x
0.0
0.1
0.2
0.3
0.4
0
2
4
6
8

Measures of Central Tendency

Mean

- The mean of a set of experimental data x_i, $i \in \{ 1, 2, \ldots, N \}$ is

$$\bar{x} = \frac{1}{N} \sum x_i$$

- The mean of the parent population is defined as the limit

$$\mu = \lim_{N \to \infty} \left(\frac{1}{N} \sum x_i \right)$$

Notation

- Parent distribution: Greek, e.g., μ, σ
- Sample distribution: Latin, $\bar{x}, \quad s$
  - To determine properties of the parent distribution assume that the properties of the sample distribution tend to those of the parent as *N* tends to infinity. Never mind that $N \rightarrow \infty$ is impractical.

Some Notation

- If we make N measurements, x_1, x_2, x_3, etc. the sum of these measurements is

$$\sum_{i=1}^{N} x_i = x_1 + x_2 + x_3 + ... + x_N$$

- Typically, we use the shorthand

$$\sum_{i=1}^{N} x_i = \sum x_i$$

Median

- The median of the parent population $\mu_{1/2}$ is the value for which half of $x_i < \mu_{1/2}$

$$P(x_i < \mu_{1/2}) = P(x_i \geq \mu_{1/2}) = 1/2$$

- The median cuts the area under the probability distribution in half

Median vs. Mean

- For a normal distribution, the average is the most efficient estimator of the mean
  - No other unbiased statistic for estimating μ has smaller variance
  - The ratio of the variance of the mean to the variance of the median for a normal distribution is $2(N-1)/(\pi N)$, where N is the sample size. As $N \rightarrow \infty$ the efficiency $\rightarrow 2/\pi \approx 0.637$

Mode

- The mode is the most probable value drawn from the parent distribution
  - The mode is the most likely value to occur in an experiment
  - For a symmetrical distribution the mean, median and mode are all the same
- An empirical relationship which appears to hold for most unimodal PDFs is

$$\mu - \mu_{\max} = 3\left(\mu - \mu_{1/2}\right)$$

Measures of Spread

Deviation

- The deviation, d_i , of a measurement, x_i , from the mean is defined as

$$d_i = x_i - \mu$$

- If μ is the true mean value the deviation is the error in x_i

Mean Deviation

- The mean deviation vanishes!
  - Evident from the definition

$$\lim_{N\to\infty} \bar{d} = \lim_{N\to\infty}\left[\frac{1}{N}\sum(x_i-\mu)\right]$$

$$= \underbrace{\lim_{N\to\infty}\left[\frac{1}{N}\sum x_i\right]}_{\mu} - \mu$$

$$= 0$$

Mean Square Deviation

- The mean square deviation is easy to use analytically & justified theoretically

$$\sigma^2 = \lim_{N\to\infty}\left[\frac{1}{N}\sum\left(x_i-\mu\right)^2\right]$$
$$= \lim_{N\to\infty}\left[\frac{1}{N}\sum x_i^2\right]-\mu^2$$

- σ^2 is also known as the ***variance***
  - Derive this expression for yourself
  - To compute σ^2 we must know μ

Standard Deviation

- A practical ***estimate*** of the standard deviation is the ***standard deviation***

$$s = \sqrt{\frac{1}{N}\sum\left(x_i - \bar{x}\right)^2}$$

- The standard deviation underestimates the dispersion of the parent population

Sample Standard Deviation

- A better ***estimate*** of σ is the ***sample standard deviation***

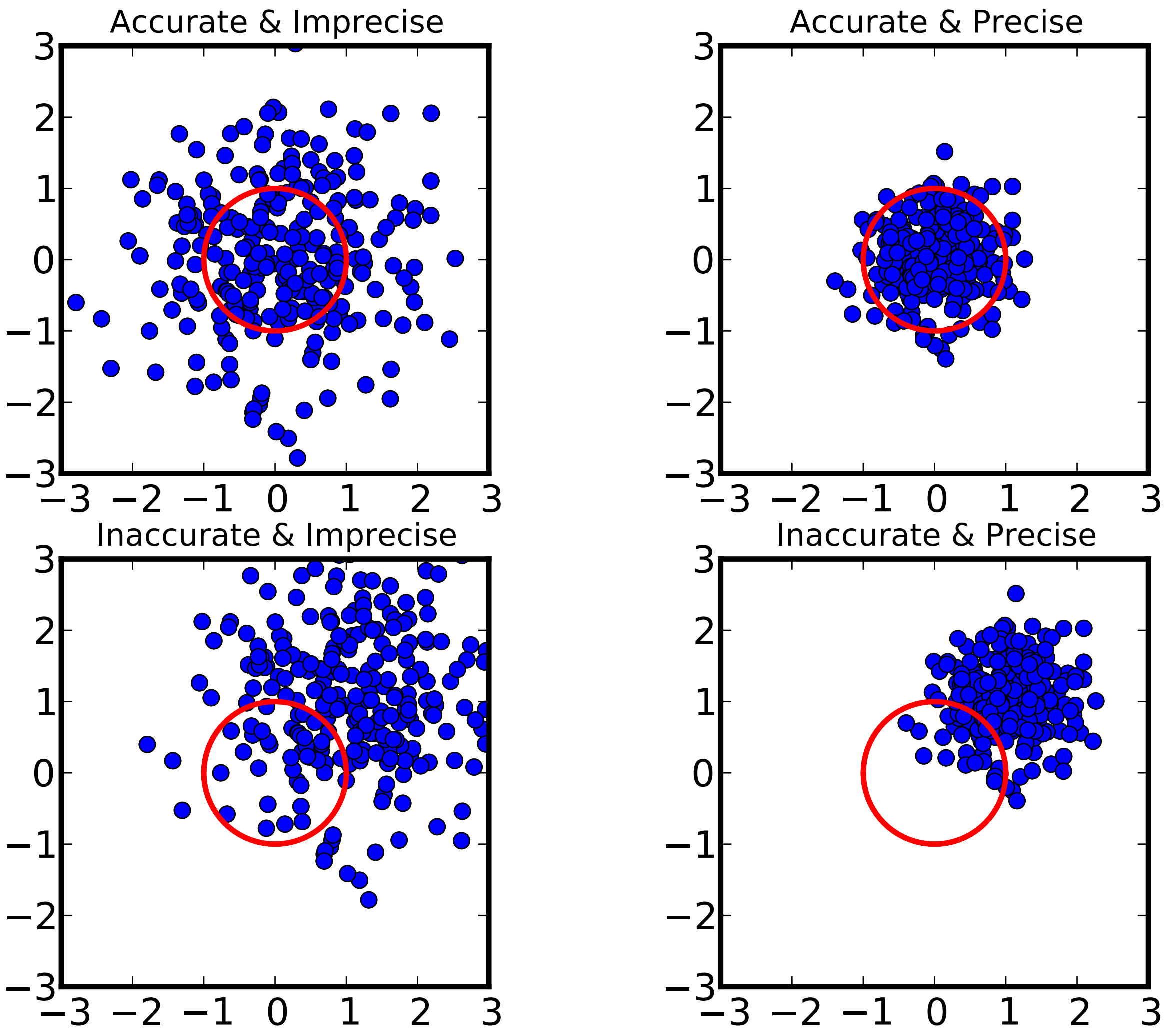
$$s = \sqrt{\frac{1}{N-1}\sum\left(x_i - \bar{x}\right)^2}$$

- "Bessel's correction factor" (*N*-1) is used instead of *N* to because the mean must be derived from the data
- s^2 is an unbiased estimator of σ^2

Significance

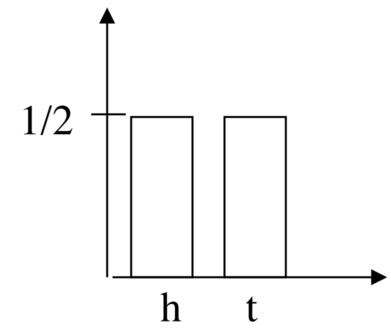
- The mean of the sample is the best estimate of the mean of the parent distribution
  - The standard deviation, s, is characteristic of the uncertainties associated with attempts to measure μ
  - But what is the uncertainty in μ?
- To answer these questions we need probability distributions…

Accuracy vs. Precision



Binomial Distribution

- Suppose we have two possible outcomes with probability *p* and *q* = 1-*p*
  - e.g., toss a fair coin, *p* = 1/2, *q* = 1/2

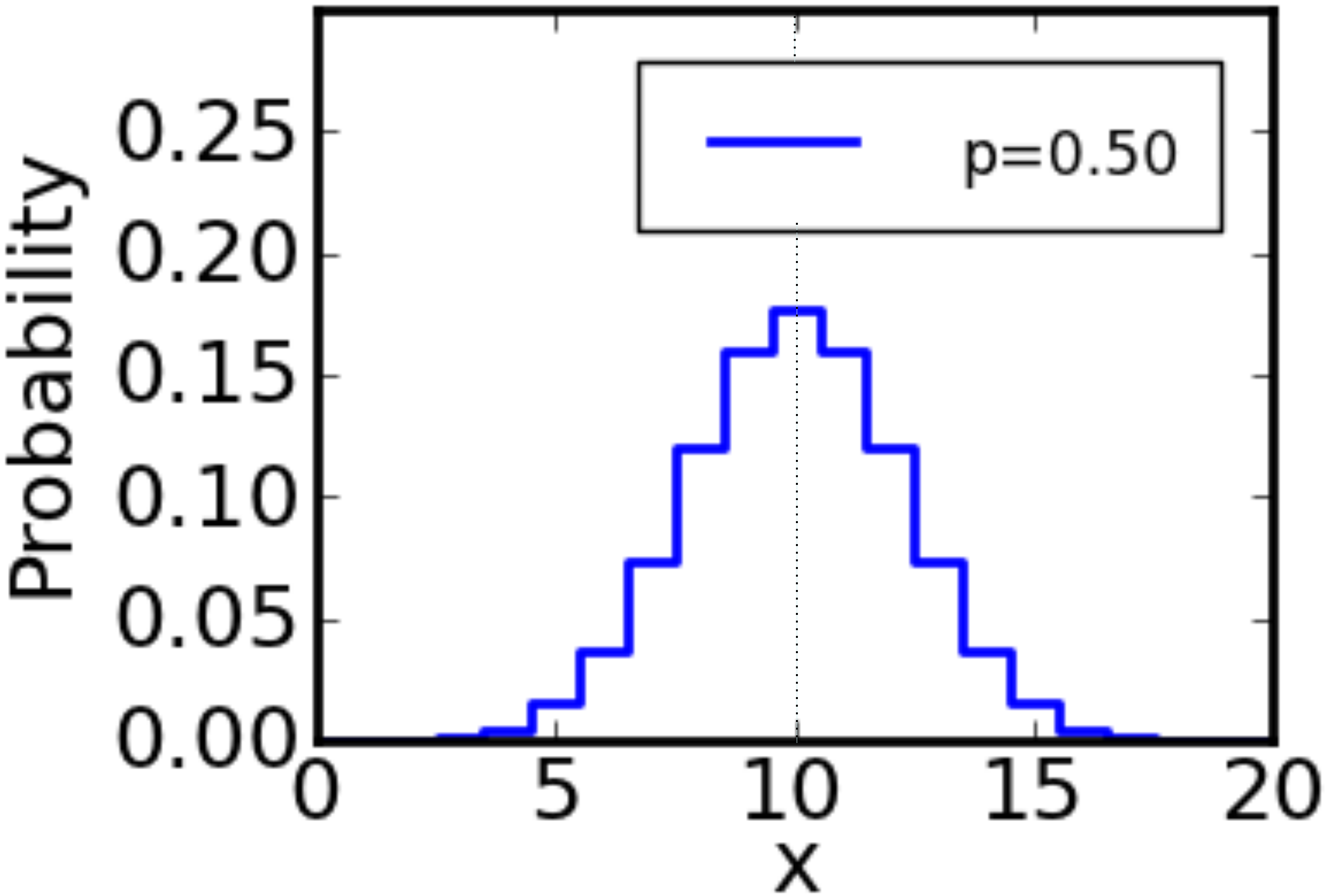


- If we flip *n* coins what is the probability of getting *x* heads?
  - Answer is given by the *Binomial Distribution*

$$P(x;n,p) = C(n,x)\,p^x q^{n-x}$$

  - *C*(*n*, *x)* or *"n* choose *x"* is the number of combinations of *n* items taken *x* at a time = *n*!/[*x*!(*n-x)!]*

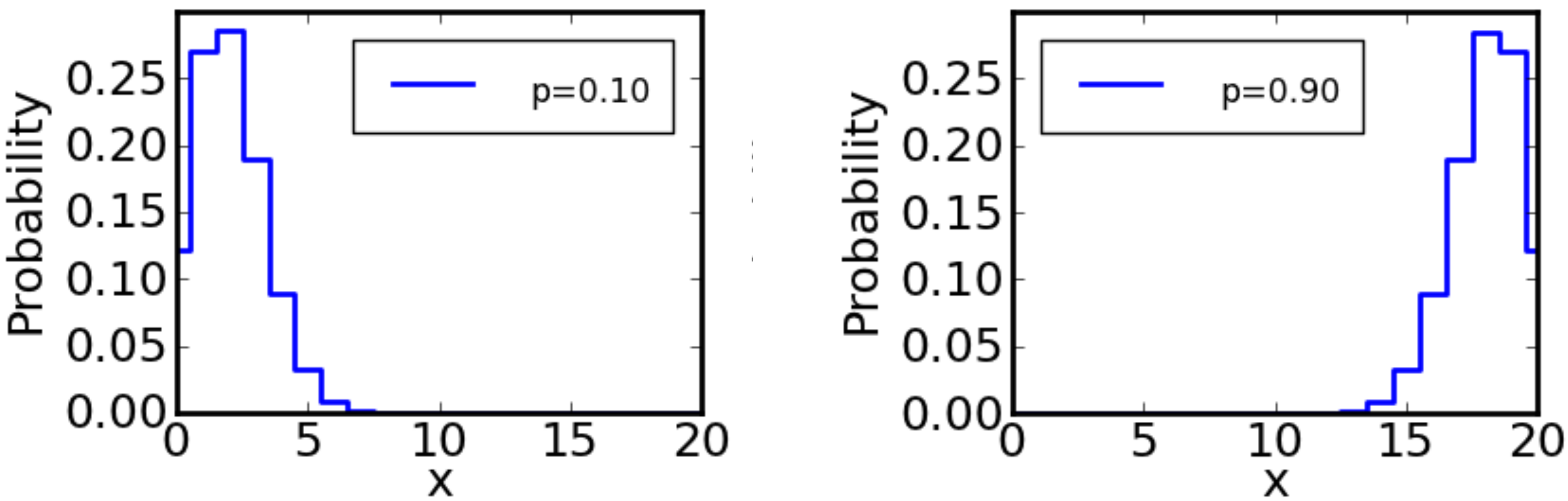
Binomial Distribution for 20 Fair Coins (n=20)



n=20

Number of heads

Binomial Distribution for Unfair or Biased Coins



Number of heads

Expectation Values

- Define μ and σ in terms of the parent probability distribution $P(x)$
- The mean, μ, is the expectation value of some quantity x

$$\langle x \rangle$$

- The variance, σ^2, is the expectation value of the deviation squared

$$\langle (x-\mu)^2 \rangle$$

Expectation Values

- For a discrete distribution, *N*, observations with *n* distinct outcomes

$$\mu = \underset{N\to\infty}{Lim}\frac{1}{N}\sum_{i=1}^{N} x_i$$

$$= \underset{N\to\infty}{Lim}\frac{1}{N}\sum_{j=1}^{n} x_j n_{x_j} \quad \text{each } x_j \text{ is a unique value}$$

$$= \underset{N\to\infty}{Lim}\frac{1}{N}\sum_{j=1}^{n} x_j NP(x_j)$$

$$= \underset{N\to\infty}{Lim}\sum_{j=1}^{n} x_j P(x_j)$$

Expectation Values

- For a discrete distribution, *N*, observations and *n* distinct outcomes

$$\sigma^2 = \underset{N\to\infty}{Lim}\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$$

$$= \underset{N\to\infty}{Lim}\frac{1}{N}\sum_{j=1}^{n}(x_j-\mu)^2 NP(x_j)$$

$$= \underset{N\to\infty}{Lim}\sum_{j=1}^{n}\left[(x_j-\mu)^2 P(x_j)\right]$$

Expectation values

- The expectation value of any continuous function of *x*

$$\langle f(x) \rangle = \int_{-\infty}^{\infty} f(x) P(x) dx$$

$$\mu = \int_{-\infty}^{\infty} x P(x) dx$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 P(x) dx$$

where $\int_{-\infty}^{\infty} P(x) dx = 1$

Binomial Distribution

- The expectation value

$$\mu = \sum_{x=0}^{n} x P(x; n, p)$$

$$= \sum_{x=0}^{n} x\, C(n, x)\, p^{x} q^{n-x}$$

$$= \sum_{x=0}^{n} \left[x \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x} \right] = np$$

Binomial Distribution

- The expectation value

$$\sigma^2 = \sum_{x=0}^{n} (x-\mu)^2 P(x;n,p)$$

$$= \sum_{x=0}^{n} (x-np)^2 C(n,x) p^x q^{n-x}$$

$$= \sum_{x=0}^{n} \left[(x-np)^2 \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \right] = np(1-p)$$

Poisson Distribution

- The Poisson distribution is the limit of the Binomial distribution when $\mu << n$ because p is small
  - The binomial distribution describes the probability $P(x; n, p)$ of observing x events per unit time out of n possible events
  - Usually we don't know n or p but we do know μ

Poisson Distribution

- Suppose *p* << 1 then *x* << *n*

$$P(x;n,p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$\frac{n!}{(n-x)!} = n(n-1)(n-2)...(n-x+2)(n-x+1)$$

$$\approx n^x \quad \text{when n >> x}$$

$$\frac{n!}{(n-x)!} p^x \approx (np)^x = \mu^x$$

$$(1-p)^{n-x} = (1-p)^{-x}(1-p)^n \approx 1 \times (1-p)^n \text{ since p << 1}$$

$$\underset{p\to 0}{Lim}(1-p)^n = \underset{p\to 0}{Lim}(1-p)^{\mu/p} = \left[\underset{p\to 0}{Lim}(1-p)^{1/p}\right]^{\mu} = \left(e^{-1}\right)^{\mu} = e^{-\mu}$$

$$P(x,\mu) = \frac{\mu^x}{x!} e^{-\mu}$$

Poisson Distribution

- The expectation value of *x* is

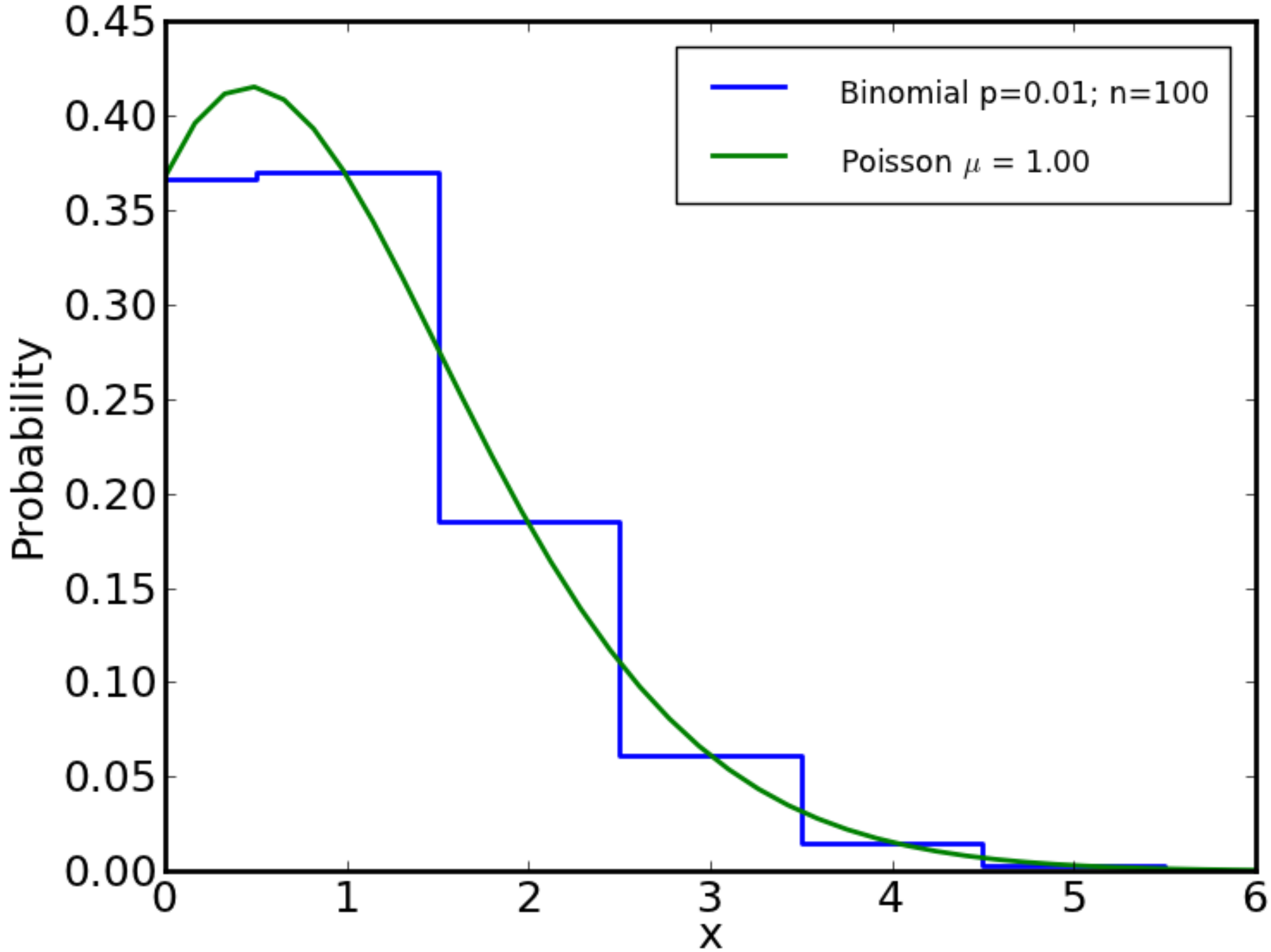
$$\langle x \rangle = \sum_{x=0}^{\infty} x P(x,\mu) = \sum_{x=0}^{\infty} x \frac{\mu^x}{x!} e^{-\mu} = \mu$$

- Expectation value of $(x\text{-}\mu)^2$

$$\sigma^2 = \left\langle \left(x - \mu \right)^2 \right\rangle = \sum_{x=0}^{\infty} (x-\mu)^2 \frac{\mu^x}{x!} e^{-\mu} = \mu$$

  – How can this be dimensionally correct?

Poisson & Binomial

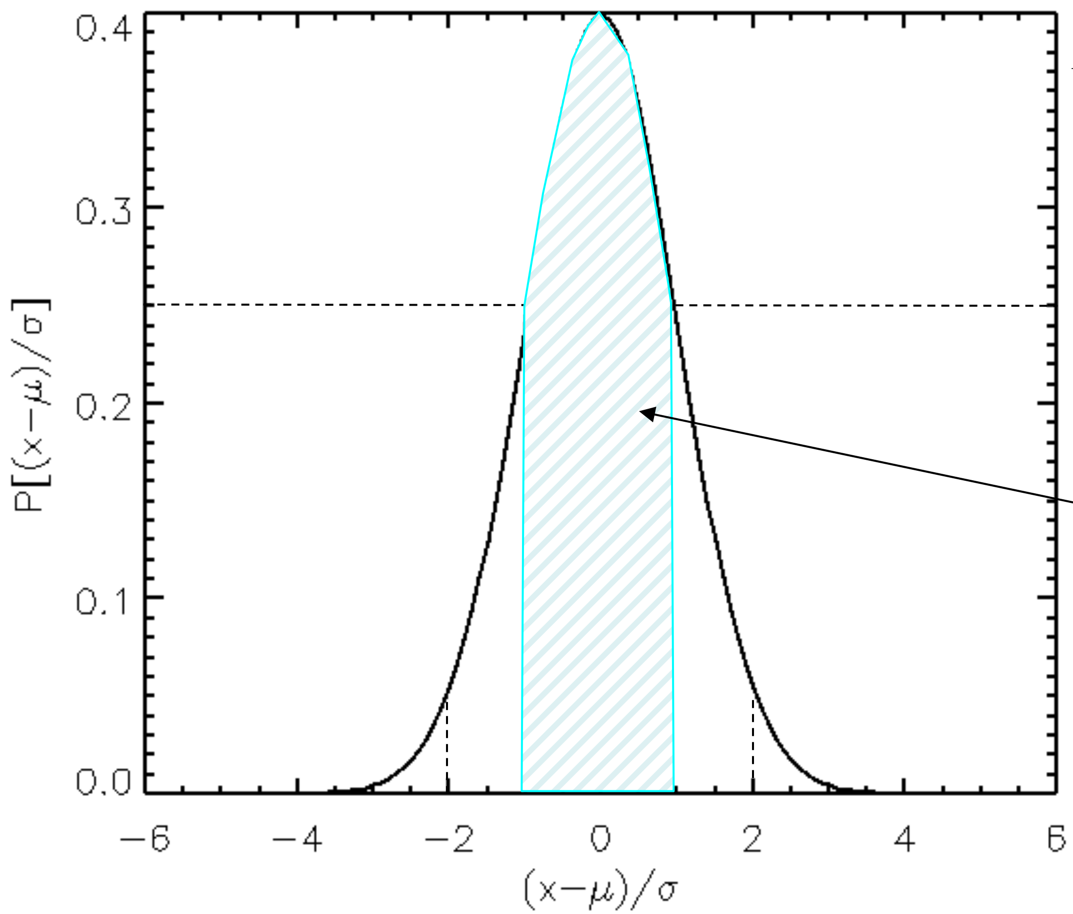


Gaussian or Normal Distribution

- The Gaussian distribution is an approximation to the binomial distribution for large *n* and large *np*

$$P(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}(x-\mu)^2\Big/\sigma^2\right]$$

Gaussian or Normal Distribution



$$P(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$$

$$\frac{1}{\sqrt{2\pi}}\int_{-1}^{1}e^{-\frac{1}{2}x^2}\,dx=0.683$$

+/- 1σ: 68.3%
+/- 2σ: 95.5%
+/- 3σ: 99.7%

Combining Observations & Error Propagation

Combining Two Observations

- Suppose I have to measure a length > 1 meter with a single meter stick.
  - I have to combine two measurement
- Two sets of measurements, a_i , and b_i
  - A derived quantity $c_i = a_i + b_i$
  - What is the relation between the means and standard deviations of a_i and b_i and c_i
  - Suppose we have the same number of observations N of a_i and b_i

Combining Two Observations

$$N = N_a = N_b$$

$$\bar{a} = \frac{1}{N}\sum a_i \quad \bar{b} = \frac{1}{N}\sum b_i$$

$$\bar{c} = \frac{1}{N}\sum c_i \quad s_c^2 = \frac{1}{N-1}\sum (c_i - \bar{c})^2$$

$$c_i = a_i + b_i$$

$$\bar{c} = \frac{1}{N}\sum (a_i + b_i) = \frac{1}{N}\sum a_i + \frac{1}{N}\sum b_i$$

$$\boxed{= \bar{a} + \bar{b}}$$

Combining Two Observations

$$s_c^2 = \tfrac{1}{N-1}\sum\left(c_i - \overline{c}\right)^2, \quad \overline{c} = \overline{a} + \overline{b}$$

$$s_c^2 = \tfrac{1}{N-1}\sum\left[a_i + b_i - \left(\overline{a} + \overline{b}\right)\right]^2$$

$$= \tfrac{1}{N-1}\sum\left[\left(a_i + b_i\right)^2 - 2\left(a_i + b_i\right)\left(\overline{a} + \overline{b}\right) + \left(\overline{a} + \overline{b}\right)^2\right]$$

$$= \tfrac{1}{N-1}\sum\left[a_i^2 + b_i^2 + 2a_ib_i - 2\left(a_i\overline{a} + a_i\overline{b} + b_i\overline{a} + b_i\overline{b}\right) + (\overline{a})^2 + 2\overline{a}\overline{b} + \left(\overline{b}\right)^2\right]$$

$$= \tfrac{N}{N-1}\overline{a^2} + \tfrac{N}{N-1}\overline{b^2} + \tfrac{2}{N-1}\sum a_ib_i - \tfrac{N}{N-1}(\overline{a})^2 - \tfrac{2N}{N-1}\overline{a}\overline{b} - \tfrac{N}{N-1}\left(\overline{b}\right)^2$$

Combining Two Observations

$$s_c^2 = \tfrac{1}{N-1}\sum\left(c_i - \overline{c}\right)^2, \quad \overline{c} = \overline{a} + \overline{b}$$

$$= \tfrac{N}{N-1}\overline{a^2} + \tfrac{N}{N-1}\overline{b^2} + \tfrac{2}{N-1}\sum a_i b_i - \tfrac{N}{N-1}(\overline{a})^2 - \tfrac{2N}{N-1}\overline{a}\overline{b} - \tfrac{N}{N-1}\left(\overline{b}\right)^2$$

$$= \tfrac{N}{N-1}\left[\overline{a^2} - (\overline{a})^2\right] + \tfrac{N}{N-1}\left[\overline{b^2} - \left(\overline{b}\right)^2\right] + \underbrace{\tfrac{2}{N-1}\sum a_i b_i - \tfrac{2N}{N-1}\overline{a}\overline{b}}_{\frac{2}{N-1}\sum\left(a_i - \overline{a}\right)\left(b_i - \overline{b}\right)}$$

$$= \underbrace{\tfrac{N}{N-1}\left[\overline{a^2} - (\overline{a})^2\right]}_{s_a^2} + \underbrace{\tfrac{N}{N-1}\left[\overline{b^2} - \left(\overline{b}\right)^2\right]}_{s_b^2} + \underbrace{\tfrac{2N}{N-1}\left(\overline{ab} - \overline{a}\,\overline{b}\right)}_{2s_{ab}}$$

$$\boxed{s_c^2 = s_a^2 + s_b^2 + 2s_{ab}}$$

- The term s_{ab} is the covariance
  - Murphy's law factor
  - s_{ab} can be negative, zero or positive

Combining Two Uncorrelated Observations

- When *a* and *b* are uncorrelated the covariance is zero

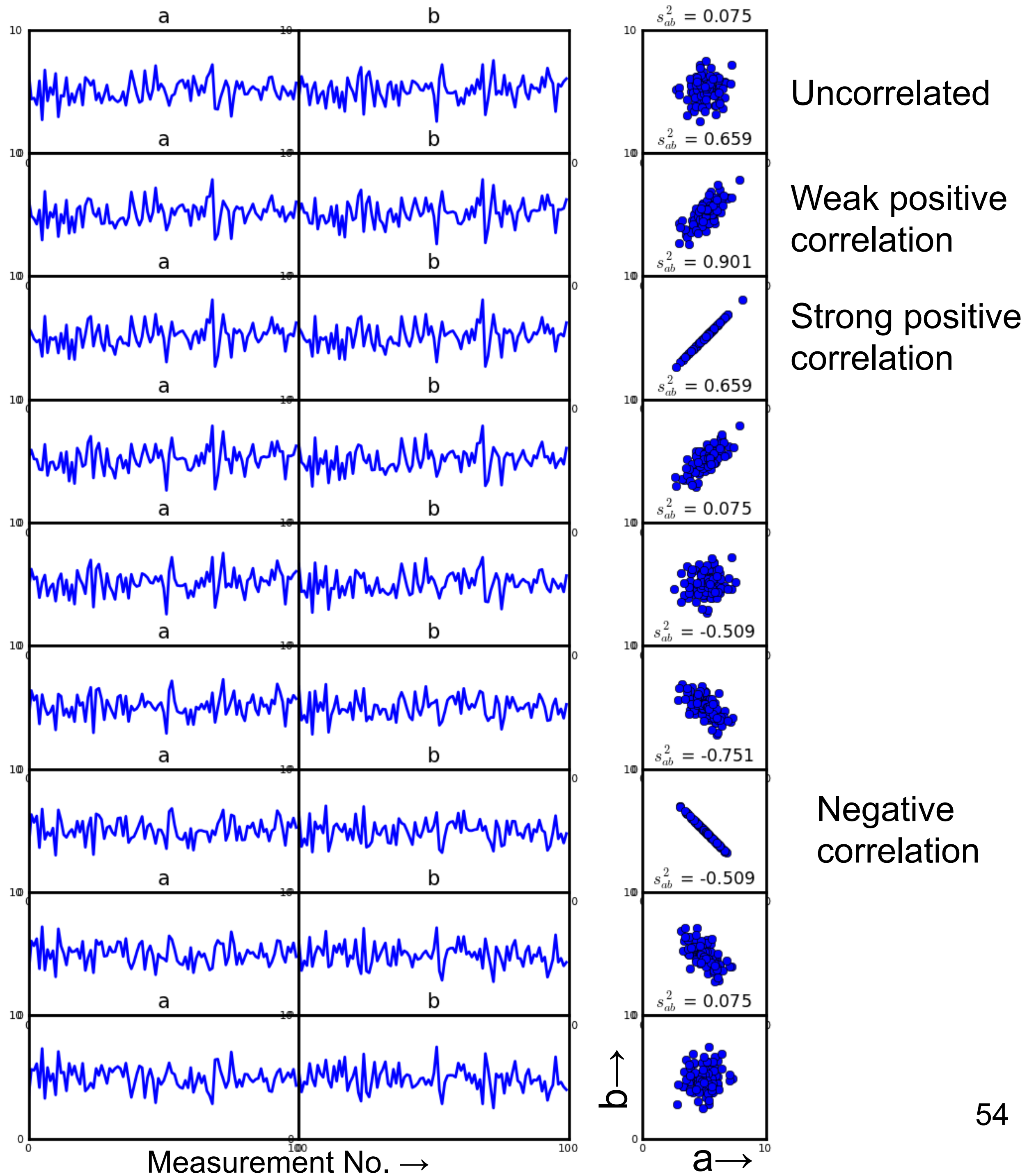
$$s_{ab} = \tfrac{1}{N-1}\sum\left(a_i - \bar{a}\right)\left(b_i - \bar{b}\right) = 0$$

$$s_c^2 = s_a^2 + s_b^2$$

  - **The variance of *c* is the sum of the variances of *a* and *b***
- This demonstrates the fundamentals of error propagation

Correlated & Uncorrelated

$$s_{ab} = \frac{2N}{N-1}\left(\overline{ab} - \overline{a}\,\overline{b}\right)$$



Propagation of Errors

- Suppose we want to determine *x* which is a function of measured quantities, *u*, *v*, etc.

$$x = f(u, v, \ldots)$$

- Assume that

$$\bar{x} = f(\bar{u}, \bar{v}, \ldots)$$

Propagation of Errors

- The uncertainty in x can be found by considering the spread of the values of x resulting from individual measurements, u_i, v_i , etc.,

$$x_i = f(u_i, v_i, \ldots)$$

- In the limit of $N \rightarrow \infty$ the variance of x

$$\sigma_x^2 = \lim_{N \to \infty} \frac{1}{N} \sum_i \left(x_i - \bar{x} \right)^2$$

Propagation of Errors

- For small errors use a Taylor approximation expand the deviation

$$x_i - \overline{x} = \left(u_i - \overline{u}\right)\frac{\partial f}{\partial u}\bigg|_{\overline{u}} + \left(v_i - \overline{v}\right)\frac{\partial f}{\partial v}\bigg|_{\overline{v}} + \ldots$$

$$\sigma_x^2 = \frac{1}{N}\sum_i \left[\left(u_i - \overline{u}\right)\frac{\partial f}{\partial u}\bigg|_{\overline{u}} + \left(v_i - \overline{v}\right)\frac{\partial f}{\partial v}\bigg|_{\overline{v}} + \ldots\right]^2$$

$$= \frac{1}{N}\sum_i \left[\left(u_i - \overline{u}\right)^2\left(\frac{\partial f}{\partial u}\right)_{\overline{u}}^2 + \left(v_i - \overline{v}\right)^2\left(\frac{\partial f}{\partial v}\right)_{\overline{v}}^2 + 2\left(u_i - \overline{u}\right)\left(v_i - \overline{v}\right)\frac{\partial f}{\partial u}\bigg|_{\overline{u}}\frac{\partial f}{\partial v}\bigg|_{\overline{v}} \ldots\right]$$

Propagation of Errors

$$\sigma_x^2 = \tfrac{1}{N}\sum_i \left[(u_i-\bar{u})^2 \left(\frac{\partial f}{\partial u}\right)_{\bar{u}}^2 + (v_i-\bar{v})^2 \left(\frac{\partial f}{\partial v}\right)_{\bar{v}}^2 + 2(u_i-\bar{u})(v_i-\bar{v}) \left.\frac{\partial f}{\partial u}\right|_{\bar{u}} \left.\frac{\partial f}{\partial v}\right|_{\bar{v}} \ldots \right]$$

$$= \tfrac{1}{N}\sum_i (u_i-\bar{u})^2 \left(\frac{\partial f}{\partial u}\right)_{\bar{u}}^2 +$$

$$\tfrac{1}{N}\sum_i (v_i-\bar{v})^2 \left(\frac{\partial f}{\partial v}\right)_{\bar{v}}^2 +$$

$$\tfrac{2}{N}\sum_i (u_i-\bar{u})(v_i-\bar{v}) \left.\frac{\partial f}{\partial u}\right|_{\bar{u}} \left.\frac{\partial f}{\partial v}\right|_{\bar{v}} + \ldots$$

$$\boxed{\sigma_x^2 = \sigma_u^2 \left(\frac{\partial f}{\partial u}\right)_{\bar{u}}^2 + \sigma_v^2 \left(\frac{\partial f}{\partial v}\right)_{\bar{v}}^2 + 2\sigma_{uv} \left.\frac{\partial f}{\partial u}\right|_{\bar{u}} \left.\frac{\partial f}{\partial v}\right|_{\bar{v}} + \ldots}$$

Examples of Error Propagation

- Suppose *a* = *b* + *c*
  - We know the exact result for combining two observations

$$\overline{a} = \overline{b} + \overline{c}$$

$$\sigma_a^2 = \sigma_b^2 + \sigma_c^2$$

  assuming that the covariance is 0
- What about *a* = *b/c*?
  - We can use the error propagation to find an estimate when the errors in *b* and *c* are small

Examples of Error Propagation

- Suppose $a = b/c$?

$$\bar{a} = \bar{b}/\bar{c}$$

and

$$\sigma_a^2 = \sigma_b^2 \left(\frac{\partial a}{\partial b} \right)_{\bar{b}}^2 + \sigma_c^2 \left(\frac{\partial a}{\partial c} \right)_{\bar{c}}^2 + 2\sigma_{bc} \left. \frac{\partial a}{\partial b} \right|_{\bar{b}} \left. \frac{\partial a}{\partial c} \right|_{\bar{c}} + \ldots$$

$$\sigma_a^2 = \sigma_b^2 \frac{1}{c^2} + \sigma_c^2 \left(\frac{b}{c^2} \right)^2$$

or

$$\frac{\sigma_a^2}{a^2} = \frac{\sigma_b^2}{b^2} + \frac{\sigma_c^2}{c^2}$$

assuming that the covariance σ_{bc} is 0

Examples of Error Propagation

- What happens when $m = -2.5 \log_{10}(F/F_0)$?
  - What is the error in m?

$$m = -2.5 \log_{10}(F/F_0)$$

and

$$\sigma_m^2 = \sigma_F^2 \left(\frac{\partial m}{\partial F} \right)_{\overline{F}}^2$$

$$\sigma_m^2 = \sigma_F^2 \left(\frac{2.5}{F \log(10)} \right)^2$$

$$\sigma_m^2 = (1.087)^2 \left(\frac{\sigma_F}{F} \right)^2$$

Error of the Mean

- Suppose I have a data set of N measurements, $x_1, x_2, x_3, \ldots$ with standard deviation s
  - *What is the standard deviation of the mean (SDOM)*

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \,?$$

  - *Error propagation shows*

$$s_{\bar{x}} = \frac{1}{\sqrt{N}}\, s$$

The standard error in the mean and SDOM are synonymous

Error of the Mean

- Suppose we have *N* measurements, x_i with uncertainties characterized by s_i

$$\bar{x} = \tfrac{1}{N}\left(x_1 + x_2 + x_3 + \ldots + x_N\right) = \tfrac{1}{N}\sum_i x_i$$

$$s_{\bar{x}}^2 = s_1^2\left(\frac{\partial \bar{x}}{\partial x_1}\right)_{\bar{x}}^2 + s_2^2\left(\frac{\partial \bar{x}}{\partial x_2}\right)_{\bar{x}}^2 + s_3^2\left(\frac{\partial \bar{x}}{\partial x_3}\right)_{\bar{x}}^2 + \ldots + s_N^2\left(\frac{\partial \bar{x}}{\partial x_N}\right)_{\bar{x}}^2$$

$$= \sum_i s_i^2\left(\frac{\partial \bar{x}}{\partial x_i}\right)_{\bar{x}}^2$$

assuming that the covariance is 0

Error of the Mean

- Suppose that the errors on all measurements are equal so $s_i = s$

$$s_{\bar{x}}^2 = \sum_i s_i^2 \left(\frac{\partial \bar{x}}{\partial x_i} \right)_{\bar{x}}^2$$

$$\frac{\partial \bar{x}}{\partial x_i} = \frac{\partial}{\partial x_i} \left(\frac{1}{N} \sum_j x_j \right) = \frac{1}{N} \qquad \text{because} \qquad \frac{\partial x_j}{\partial x_i} = \delta_{ij}$$

$$\boxed{\begin{aligned} s_{\bar{x}}^2 &= \sum_i s^2 \left(\frac{1}{N} \right)^2 \\ &= \frac{s^2}{N} \end{aligned}}$$

Error Propagation II

- The scheme for error propagation uses a Taylor *approximation*
  - Valid only for small errors
  - Consider the example where $y = x^2$

Our Taylor-series based approximation says

$$\sigma_y^2 = \left(\frac{\partial y}{\partial x}\right)_{\bar{x}}^2 \sigma_x^2$$

$$\sigma_y^2 = 4\bar{x}^2\sigma_x^2$$

Error Propagation II

- The correct method requires that we know the PDF for *y*
  - Suppose we know the PDF for *x*, *f*(*x*) what is the PDF for *y*, *g*(*y*)?

$$\int_{-\infty}^{\infty} g(y)\,dy = \int_{-\infty}^{\infty} f(x)\,dx$$

The only way this can be true in general if

$$g(y)\,dy = f(x)\,dx$$

Error Propagation II

- For the case where $y = x^2$

Find the mean of y: $\quad \bar{y} = \int_{-\infty}^{\infty} y g(y) dy = \int_{-\infty}^{\infty} x^2 f(x) dx$, i.e., $\bar{y} = \langle x^2 \rangle$

Now $\sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2$ or $\bar{y} = \bar{x}^2 + \sigma_x^2$ and

$$\sigma_y^2 = \int_{-\infty}^{\infty} (y - \bar{y})^2 g(y) dy = \int_{-\infty}^{\infty} y^2 g(y) dy - (\bar{y})^2 = \int_{-\infty}^{\infty} x^4 f(x) dx - (\bar{y})^2 = \langle x^4 \rangle - \langle x^2 \rangle^2$$

$\langle x^4 \rangle = \bar{x}^4 + 6\bar{x}^2\sigma_x^2 + 3\sigma_x^2 \qquad$ when $f(x)$ is Gaussian

So the accruate answer is $\quad \sigma_y^2 = 4\bar{x}^2\sigma_x^2 + 2\sigma_x^4$

But if $f(x)$ is Poisson we get a different answer because

$\langle x^2 \rangle = \bar{x}^2 + \bar{x}, \; \langle x^4 \rangle = \bar{x}^4 + 6\bar{x}^3 + 7\bar{x}^2 + \bar{x} \qquad$ and $\sigma_y^2 = \bar{x}\left(4\bar{x}^2 + 6\bar{x} + 1\right)$

Error Propagation II

- Comparison of error propagation methods for $y = x^2$

$y = x^2$	Gaussian	Poisson
μ_x	1	1
σ_x	1	1
μ_y: y> = x^2	$\mu^2 + \sigma^2 = 2$	$\mu^2 + \mu = 2$
Taylor: $\sigma_y^2 = 4\ \mu^2\ \sigma_x^2$	4	4
Exact: σ_y^2	$4\ \mu_x^2\ \sigma_x^2 + 2\sigma_x^4 = 6$	$\mu_x(4\mu_x^2 + 6\mu_x + 1) = 11$